

Spatio-Temporal-Spectral Hierarchical Graph Convolutional Network With Semisupervised Active Learning for Patient-Specific Seizure Prediction

Yang Li¹, Yu Liu, Yu-Zhu Guo¹, Xiao-Feng Liao¹, Fellow, IEEE, Bin Hu¹, Senior Member, IEEE, and Tao Yu²

Abstract—Graph theory analysis using electroencephalogram (EEG) signals is currently an advanced technique for seizure prediction. Recent deep learning approaches, which fail to fully explore both the characterizations in EEGs themselves and correlations among different electrodes simultaneously, generally neglect the spatial or temporal dependencies in an epileptic brain and, thus, produce suboptimal seizure prediction performance consequently. To tackle this issue, in this article, a patient-specific EEG seizure predictor is proposed by using a novel spatio-temporal-spectral hierarchical graph convolutional network with an active preictal interval learning scheme (STS-HGCN-AL). Specifically, since the epileptic activities in different brain regions may be of different frequencies, the proposed STS-HGCN-AL framework first infers a hierarchical graph to concurrently characterize an epileptic cortex under different rhythms, whose temporal dependencies and spatial couplings are extracted by a spectral-temporal convolutional neural network and a variant self-gating mechanism, respectively. Critical intrarhythm spatiotemporal properties are then captured and integrated jointly and further mapped to the final recognition results by using a hierarchical graph convolutional network. Particularly, since the preictal transition may be diverse from seconds to hours prior to a seizure onset among different patients, our STS-HGCN-AL scheme estimates an optimal preictal interval patient dependently via a semisupervised active learning strategy, which further enhances the robustness of the proposed patient-specific EEG seizure predictor. Competitive experimental results validate the efficacy of the proposed method in extracting critical preictal biomarkers, indicating its promising abilities in automatic seizure prediction.

Manuscript received September 8, 2020; revised February 4, 2021; accepted March 26, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant U1809209 and Grant 61671042; in part by the Beijing Natural Science Foundation under Grant L182015; in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LSZ19F020001; and in part by the Major Project of Wenzhou Natural Science Foundation under Grant ZY2019020. This article was recommended by Associate Editor Y. Jin. (*Corresponding author: Yu Liu*)

Yang Li is with the Department of Automation Sciences and Electrical Engineering, Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100083, China (e-mail: liyang@buaa.edu.cn).

Yu Liu and Yu-Zhu Guo are with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100083, China (e-mail: sy1803113@buaa.edu.cn; yuzhuguo@buaa.edu.cn).

Xiao-Feng Liao is with the College of Computer Science, Chongqing University, Chongqing 400715, China (e-mail: xfliao@cqu.edu.cn).

Bin Hu is with the School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China (e-mail: bh@lzu.edu.cn).

Tao Yu is with the Beijing Institute of Functional Neurosurgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China (e-mail: yutaoly@sina.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2021.3071860>.

Digital Object Identifier 10.1109/TCYB.2021.3071860

Index Terms—Active learning, electroencephalogram (EEG), graph convolutional network (GCN), seizure prediction, spatio-temporal-spectral dependencies.

I. INTRODUCTION

EPILEPSY is one of the most common neurological disorders and affects over 40 million people worldwide [1], [2]. Generally, a system capable of predicting the occurrence of seizures from electroencephalogram (EEG) signals can significantly improve the therapeutic possibilities for epilepsy patients [3]. However, it is difficult to capture useful preictal biomarkers for efficient seizure prediction due to the complex spatiotemporal dynamics in an epileptic brain during the initiation of seizures [2], [4].

In order to realize automatic EEG seizure prediction, many deep learning-based approaches have been studied [5], [6]. Most of them aim to exploit the spatiotemporal evolutions in EEGs for predicting seizures [2], [7] since seizures occur as a result of both spatial and temporal evolutions in an epileptic brain [2]. Unfortunately, conventional deep-learning architectures [8], [9], the convolutional neural network (CNN) for instance, can only capture short-range spatial correlations among EEG channels due to the regular and local receptive field of convolution operators [10]. However, recent studies reported that abnormal interactions in an epileptic brain involve not only the seizure-onset focal but also long-range correlations [4]. Thus, recent advanced deep learning strategies, which can capture irregular spatiotemporal nature in EEGs efficiently, should be highly required for seizure prediction [4], [11].

To tackle the above-mentioned issues, the graph convolutional network (GCN) is investigated, since GCN-based architectures could capture arbitrary spatiotemporal responses from EEGs by viewing it as signals represented by graphs [10]. Specifically, EEG signals are first converted to a graph representation by associating its spatial and temporal properties with graph edges and nodes, respectively, [12]. The resulting EEG graphs can then be fed into a GCN network to capture its irregular spatiotemporal attributes [12], [13]. However, two main deficiencies exist in the above EEG graph construction. First, although intracranial EEGs measure the electrical activities within a cortical network directly, intracranial signals usually only cover a limited cortical area [14]. Thus,

in order to characterize interactions between different brain regions within a whole-brain range, recent studies generally construct EEG graphs based on scalp EEGs due to its good coverage [10], [13]. However, scalp potentials cannot reflect the activities in the underlying cortical network univocally due to its volume conduction effect [14]. Thus, graph construction from EEGs should be preceded by a source projection preprocessing to reconstruct the source cortical activities that underly the given scalp potentials, which should boost the representational abilities of the resulting EEG graphs in characterizing an epileptic cortex precisely. Second, most studies merely use handcrafted features to estimate graphs representing EEGs [12], [13]. For example, Wang *et al.* [13] used phase-locking value (PLV) to infer the spatial couplings among EEG channels, that is, graph edges. Differential entropy (DE) was adopted to realize temporal correlations in EEGs, namely, graph nodes [15]. However, priori indicators ignore the heterogeneity among epilepsy patients, which potentially degrade the generalization abilities of the resulting EEG graphs and subsequent GCN models for seizure prediction consequently [16]. Thus, an integrated GCN architecture, which is able to both infer graphs to represent the epileptic cortex patient dependently and extract its arbitrary spatiotemporal responses with graph convolutions jointly, may be more advantageous for seizure prediction.

In contrast to conventional deep learning approaches [2], [7], GCN-based networks could exploit irregular spatiotemporal evolutions in epileptic activities and, thus, may own better abilities for seizure prediction [11]. However, spatiotemporal responses regardless of spectral attributes commonly ignore the hierarchical organization of an epileptic brain, that is, different brain regions in terms of epileptogenicity may be dominated by different rhythmic activities [17] and, thus, may not be fully qualified for seizure prediction. For example, the epileptogenic zone, that is, brain regions that generate seizures, is governed by fast activities, while activities in the propagation zone, that is, regions triggered by seizures in epileptogenic zone, are usually of lower frequencies [18]. Inspired by these observations, in this article, we aim to extend the hierarchically organized GCN network to reconcile multirhythm spatiotemporal dependencies in an epileptic brain for better seizure prediction.

In addition, considering the severe variability of epileptic seizures, the preictal transition may vary from seconds to hours prior to a seizure onset [2]. As a consequence, supervised learning-based studies have to presuppose a fixed preictal interval and label training EEGs empirically to initialize model optimization [2], [7]. However, a predefined preictal interval is generally error-prone, since the actual duration of a preictal transition, that is, the exact start point of the preictal period, is still missing [6]. Moreover, an improper preictal interval may result in training EEGs with erroneous annotations, which potentially degrades the performance of a supervised seizure predictor [6]. Thus, it is more appropriate to regard the seizure prediction task as an incomplete supervision problem, that is, only a small subset of training EEGs, that is, samples that are either close enough or far enough to the related seizure onset,

can be given with labels empirically while others are unlabeled initially. Since the latest active learning strategies have proven to be an efficient solution to addressing incomplete supervision issues by assuming that there is an “oracle” that can be queried to obtain ground-truth labels for the initially unlabeled samples [19], we further propose an automatic preictal interval estimation method by integrating active learning strategies with the semisupervised learning method.

To address the aforementioned issues, in this article, we propose a novel patient-specific EEG seizure predictor, called spatio-temporal-spectral hierarchical GCN with an active preictal interval learning scheme (STS-HGCN-AL). Specifically, the STS-HGCN-AL includes three subnets. The spectral-temporal squeeze-and-excitation subnet (ST-SENet) first takes independent components (ICs) identified via IC analysis (ICA) as inputs and captures its temporal evolutions under different rhythms parallelly. Meanwhile, the sparse self-gating subnet (GATENet) estimates the strengths of spatial couplings between ICs adaptively and prunes weak connections automatically. Thus, a hierarchical graph, which is able to characterize multirhythm spatial and temporal dynamics in an epileptic cortex, is obtained. Intrarhythm spatiotemporal properties are then captured and integrated alternatively and further mapped to the final recognition results by using the hierarchical graph convolution subnet (HGCN). In addition, a semisupervised active learning strategy is nested in our STS-HGCN-AL scheme to infer the patient-specific optimal preictal interval adaptively. Experimental results on multiple patients validate that our system could well capture critical preictal biomarkers and is stronger for seizure prediction than the state-of-the-art methods.

The main contributions of our study are summarized as follows.

- 1) A novel STS-HGCN-AL scheme is proposed for automatic seizure prediction, which is capable of combating the heterogeneity among patients by inferring graphs to represent EEGs and exploring its irregular spatiotemporal responses jointly. Our model is publicly available.¹
- 2) We propose two variant graph convolutions: a) residual graph convolution (resGCN) and b) rhythm attention (rhythmAtt) units, to better capture the preictal EEG transitions by hierarchically reconciling spatiotemporal evolutions in an epileptic brain under different rhythms.
- 3) A semisupervised active learning strategy is investigated to infer the patient-specific optimal preictal interval adaptively, which further enhances the robustness of our seizure predictor.

II. PRELIMINARY

Typically, a directed graph \mathcal{G} can be represented as $\mathcal{G} = \{\mathcal{V}, \mathcal{X}, \mathcal{A}\}$, where \mathcal{V} is the vertex set with $|\mathcal{V}| = n$ nodes, $\mathcal{X} \in R^{n \times c}$ is the node attributes, whose i th row $x_i \in R^c$ denotes the c -channel features indexed by node i , and $\mathcal{A} \in R^{n \times n}$ is the adjacent matrix, whose (i, j) th entry indicates connections from node j to i . Convolving on graph \mathcal{G} at the k th layer can then be

¹Our model will be available at <https://github.com/YangLibuaa/STS-HGCN-AL> or <https://github.com/YuLiu-web/STS-HGCN-AL>.

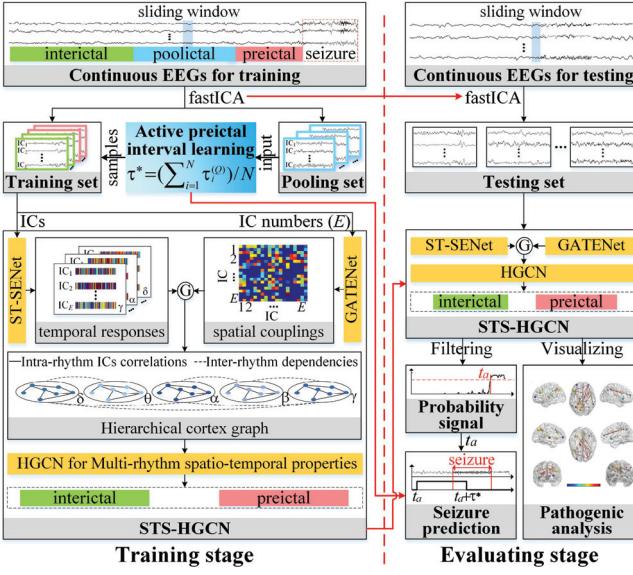


Fig. 1. Flowchart of our proposed STS-HGCN-AL for EEG seizure prediction, where the STS-HGCN architecture, including subnets ST-SENet, GATENet, and HGNC, is adopted to capture multirhythm spatiotemporal dependencies in an epileptic cortex jointly. In addition, a semisupervised active learning strategy is nested in the training stage of the STS-HGCN to infer the patient-specific optimal preictal interval adaptively. Finally, a practical seizure warning system is obtained by postprocessing optimal STS-HGCN-AL resulted from the training stage via a persistent warning scheme.

generally defined as the following aggregating and updating scheme [20]:

$$h_i^{(k+1)} = \eta^{(k)}\left(h_i^{(k)}, \varphi^{(k)}\left(h_i^{(k)}, \left\{\left(h_j^{(k)}, A_{i,j}\right) | j \in \mathcal{N}_i\right\}\right)\right) \quad (1)$$

where $h_i^{(k)}$ is the features on node i at the k th layer; $h_i^{(0)} = x_i$, \mathcal{N}_i is the neighboring set of node i ; and $\varphi^{(k)}(\cdot)$ and $\eta^{(k)}(\cdot)$ are node feature aggregator and updater at the k th layer, respectively. Thus, at the k th layer, each node in \mathcal{G} adopts aggregator $\varphi^{(k)}(\cdot)$ to compile information from its neighbors and uses updater $\eta^{(k)}(\cdot)$ to update its own attributes [20]. Iteratively, the local structures of \mathcal{G} are gathered as the resulting node attributes.

III. METHODOLOGY

The proposed patient-specific EEG seizure predictor, which is based on the STS-HGCN-AL framework, is outlined in Fig. 1 and summarized as follows: 1) apply the STS-HGCN architecture, which includes three basic subnets, that is: a) the ST-SENet; b) the GATENet; and c) the HGNC, to capture multirhythm spatiotemporal evolutions in an epileptic cortex and generate recognition results; 2) infer the patient-specific optimal preictal interval adaptively via a semisupervised active learning strategy, which is efficient to combat the variability of seizures; and 3) translate obtained optimal STS-HGCN-AL from training stage into a practical seizure warning system, which considers the interactions with patients carefully. The details of each step are given in the following sections.

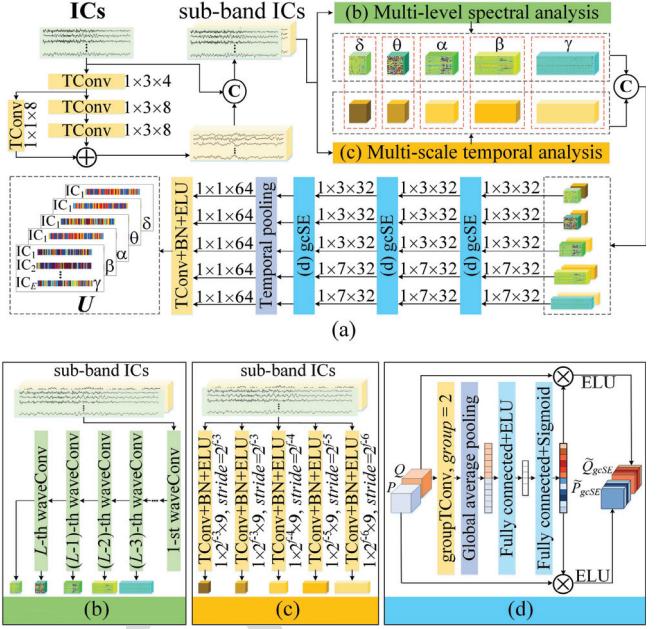


Fig. 2. (a) Diagram of the ST-SENet, which uses temporal embedding block, (b) multilevel spectral analysis, (c) multiscale temporal analysis, and stacked (d) gcSE units to capture intrarhythm ICs temporal responses $U = [u_\gamma, u_\beta, u_\alpha, u_\theta, u_\delta]$ with $u_i \in R^{E \times 64}$, where the j th row of u_i means temporal evolutions in the j th IC under the i th rhythm uniquely.

A. Spatio-Temporal-Spectral Hierarchical Graph Convolutional Network Architecture

In this section, we employ the STS-HGCN architecture, including subnets ST-SENet, GATENet, and HGNC, to infer a hierarchical graph of the epileptic cortex under different rhythms and capture its multirhythm spatiotemporal attributes. We then use matrix $X \in R^{E \times T}$ to represent a t -second EEG epoch with E channels and T sampling indices at a sampling rate of 2^f Hz, where $t = T/2^f$.

1) *Design of the ST-SENet:* In order to characterize the temporal evolutions in an epileptic cortex directly, the ST-SENet takes ICs identified via ICA as inputs, rather than scalp EEGs. This is mainly inspired by the previous findings that ICA methods could map mixed scalp EEGs to mutually independent ICs, each of which arises from local field activities in a certain cortical area uniquely [21]. Thus, ICs identified by ICA reflect cortical voltage activities precisely and, thus, facilitate the ST-SENet to better capture temporal dynamics in an epileptic brain. Specifically, we use fastICA projection to map scalp EEGs to ICs due to its better convergence properties than other ICA methods [22]. In order to ensure a fully automatic EEG seizure predictor, no manual elimination for noncortical ICs, for example, eye blink, is adopted. As a result, the ST-SENet takes a 3-D matrix with a size of $1 \times E \times T$, that is, E -channel ICs, as inputs, and captures its temporal responses under different rhythms independently. A diagram of the ST-SENet is given in Fig. 2, which includes three basic operators, that is: 1) the temporal embedding block; 2) the multilevel spectral and multiscale temporal analyses; and 3) the group convolution squeeze & excitation (gcSE) operations.

Because convolution operators essentially equate to a low-pass filter [23], the temporal embedding block, that is, successive temporal convolution and batch normalization (BN) operations, is first adopted to infer a patient-specific optimal filter-band for the subsequent analysis. Therefore, the heterogeneity among patients can be alleviated [16]. Moreover, since temporal convolutions allow no information interflow between different ICs, the i th row of the output embeddings corresponds to the sub-band responses of the i th IC uniquely, where $i = 1, 2, \dots, E$. As a result, after stacking input ICs and output embeddings with a channelwise concatenating function, the ST-SENet obtains a sub-band ICs matrix with a size of $9 \times E \times T$, which provides a subsequent network with adaptive sub-band responses of input ICs and original IC series as well.

Next, since epileptic activities in different brain regions may be of different frequencies [18], spectral feature responses, approximately corresponding to the dynamic ICs under standard physiological sub-bands δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–50 Hz), respectively, [24], [25], are obtained by the multilevel spectral analysis, that is, L -layer Daubechies order-4 (Db4) wavelet-based wavelet convolutions (waveConv). Thereby, L is equal to $f-3$, which is determined by the sampling rate 2^f Hz. Particularly, we select Db4 wavelet since previous studies reported that Db4 mother wavelet is useful for epileptiform transient detection due to its high correlation coefficients with the epileptic spike signal [26], [27]. Note that since waveConv operators realize multiresolution spectral analysis in a way analogous to the discrete wavelet transform [28], frequency boundary of the i th spectral feature response corresponding to the j th standard physiological sub-band is equal to $(f_{i,j}/2^{l+1})$, if the j th standard physiological sub-band is not δ rhythm; otherwise, $(0, f_{i,j}/2^{l+1})$, where $i = 1, 2, \dots, (E+8)$, $f_{i,j}$ is the adaptive filter-band and l is the corresponding resolution level. Thus, frequency boundaries of different spectral feature responses corresponding to the same standard physiological sub-band are slightly different from each other, since $f_{i,j}$ is learned by the above temporal embedding block in a data-driven way. As a result, interpatient variation of critical frequency boundaries, caused by age differences for example [29], may be mitigated to some extent. However, it is very difficult to yield good generalization abilities for predicting seizures based on the predefined Db4 wavelet solely, since typical patterns under the preictal transition can be diverse among different patients [11]. Thus, multiscale temporal analysis, that is, temporal convolution with trainable kernel parameters, BN and, exponential linear unit (ELU) operations, captures temporal embeddings of the dynamic ICs at different scales in a data-driven way. ELU nonlinearity is adopted due to its abilities of fast and accurate deep network learning in EEG decoding tasks [30]. As a result, five groups of intrarhythm spectral-temporal feature responses, which characterize the dynamic ICs under δ , θ , α , β , and γ rhythms, respectively, are obtained after concatenating the above mutually complementary spectral and temporal embeddings via a channelwise stacking function. More details of the multilevel spectral and multiscale temporal analyses can be found in our previous work [28].

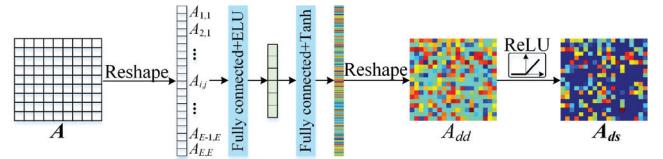


Fig. 3. Structure of the GATENet, where a sparse self-gating block is used to embed impacts of the j th IC on the i th IC in the (i, j) th entry of A_{ds} .

Finally, the ST-SENet branches into five parallel subnets, each of which takes a certain group of primary intrarhythm spectral-temporal embeddings as inputs and extracts high-level intrarhythm responses via successive gcSE units. Specifically, given a pair of intrarhythm spectral-temporal responses P and Q , the gcSE unit captures advanced spectral and temporal features \tilde{P}_{gcSE} and \tilde{Q}_{gcSE} separately by using group temporal convolution, due to the heterogeneity of multidomain features in nature. However, although no information outflow is allowed between spectral and temporal branches, both of them can be regarded as a collection of local descriptors, whose statistics are significant for the entire input ICs. Thus, next, the gcSE unit boosts cross-domain critical responses jointly via a squeeze & excitation block [28]. A detailed discussion of the gcSE unit can be found in our previous work [28]. As a result, after compacting advanced intrarhythm embeddings over time, the ST-SENet obtains coarser intrarhythm ICs temporal responses $U = [u_\gamma, u_\beta, u_\alpha, u_\theta, u_\delta]$ with $u_i \in R^{E \times 64}$, where j th row of u_i is temporal evolutions in the j th IC under the i th rhythm uniquely. Choices of the parameters in network, such as *kernel size* and *stride* in convolution operators, can be found in our previous work [28].

2) Design of the GATENet: Aiming to infer the ICs spatial couplings, the structure of the GATENet is shown in Fig. 3. Specifically, the GATENet first considers an initial adjacent matrix $A \in R^{E \times E}$ as inputs, whose (i, j) th entry denotes coupling strengths between the i th and j th ICs and is set as 1 [14]. Thus, A indicates that every IC is densely connected with each other, regardless of weight and direct. However, recent studies reported the necessity of measuring degree of directional influences between different brain regions, that is, ICs in our study, for characterizing an epileptic brain [14]. Thus, the GATENet uses a self-gating block with Tanh nonlinearity to model the directional dependencies between ICs

$$\tilde{A}_{dd} = \sigma(W_2 \delta(W_1 \tilde{A})) \quad (2)$$

where $\tilde{A} \in R^{(E \times E) \times 1}$ is vectorized from A , $W_1 \in R^{((E \times E)/r) \times (E \times E)}$ and $W_2 \in R^{(E \times E) \times ((E \times E)/r)}$ are weight matrixes, and $\delta(\cdot)$ and $\sigma(\cdot)$ are ELU and Tanh functions, respectively. Choices of the reduction ratio r will be discussed in Section IV-D. As a result, a dense adjacent matrix $A_{dd} \in R^{E \times E}$, whose (i, j) th element reflects the directional influence from j th to i th ICs, is obtained by reshaping $\tilde{A}_{dd} \in R^{(E \times E) \times 1}$ into $R^{E \times E}$. However, since we do not manually prune non-cortical ICs in the previous section, spurious couplings in A_{dd} are unavoidable. Thus, since the Tanh function in the previous self-gating unit regularizes entries of A_{dd} into $(-1, 1)$ interval, a rectified linear unit (ReLU) is adopted to prune weak ICs

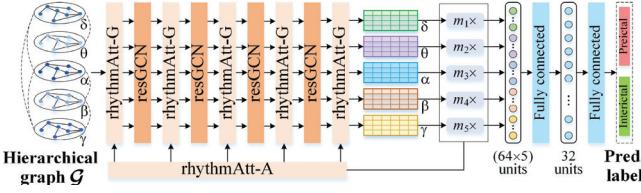


Fig. 4. Illustration of the HGCN, where multirhythm spatiotemporal responses are captured by using our resGCN and rhythmAtt units iteratively.

346 spatial correlations in A_{dd} automatically, that is, entities with
347 negative values. In other words, the combination of Tanh and
348 ReLU nonlinearities is adopted to penal weak ICs couplings
349 and result in a sparse brain network. As a result, a sparse
350 adjacent matrix $A_{ds} \in R^{E \times E}$, which is non-negative and only
351 retains significant spatial correlations in an epileptic cortex, is
352 obtained.

353 3) *Design of the HGCN*: Based on the previous intrarhythm
354 ICs temporal responses $U = [u_\gamma, u_\beta, u_\alpha, u_\theta, u_\delta]$ with $u_i \in$
355 $R^{E \times 64}$ and sparse adjacent matrix $A_{ds} \in R^{E \times E}$, that is, ICs spa-
356 tial correlations, a hierarchical cortex graph is constructed to
357 concurrently characterize the spatial and temporal dependen-
358 cies in an epileptic cortex under different rhythms. Specifically,
359 intrarhythm graph instances g_i are first obtained, where g_i with
360 $i = 1, 2, \dots, 5$ correspond to graph representations of the cor-
361 tical network under $\delta, \theta, \alpha, \beta$, and γ rhythms, respectively.
362 For i th rhythm, g_i is defined as $g_i = \{C_i, u_i, A_{ds}\}$, where C_i
363 is the vertex set with $|C_i| = E$ nodes, node attributes are
364 ICs temporal responses under the i th rhythm $u_i \in R^{E \times 64}$,
365 and connections between nodes are determined by the sparse
366 adjacent matrix $A_{ds} \in R^{E \times E}$. Inspired by the previous find-
367 ings that interrhythm energy shifting may be a biomarker of
368 the seizure stages changing [17], intrarhythm graphs g_i are
369 further interconnected to facilitate inter-rhythm information
370 outflow, thus forming the hierarchical graph G . Therefore,
371 it is defined as $G = \{\mathcal{V}, \{g_i, i = 1, 2, \dots, 5\}, L\}$, where \mathcal{V}
372 is the vertex set with $|\mathcal{V}| = 5$ nodes, intrarhythm graphs
373 $\{g_i, i = 1, 2, \dots, 5\}$ are node attributes, and $L \in R^{5 \times 5}$ is intern-
374 ode connections, that is, inter-rhythm dependencies, and will
375 be learned by the rhythmAtt unit subsequently. As a result,
376 based on the hierarchical graph, the HGNCN could capture
377 multirhythm spatiotemporal responses jointly by using our res-
378 GCN and rhythmAtt units iteratively. A diagram of the HGNCN
379 is shown in Fig. 4.

380 Motivated by the independence between different rhythmic
381 activities, intrarhythm spatiotemporal properties are captured
382 parallelly. For this purpose, a variant graph convolution, called
383 resGCN, is defined on the intrarhythm graph $g_i = \{C_i, u_i, A_{ds}\}$
384 by integrating graph convolutions with residual learning

$$385 H^{(i,k+1)} = \delta_2 \left(D^{-1} A_{ds} \delta_1 \left(H^{(i,k)} W_1^{(i,k)} \right) W_2^{(i,k)} + \sum_{j=0}^{k-1} H^{(i,j)} \right) \quad (3)$$

386 where $k = 0, 1, \dots, K - 1$ is the number of layers, $D^{ii} =$
387 $\sum_j A_{ds}^{ij}$ is the degree matrix of A_{ds} , $H^{(i,k)} \in R^{E \times N}$ is the
388 activation at the k th layer, $H^{(i,0)} = u_i$, $W_1^{(i,k)} \in R^{N \times N}$
389 and $W_2^{(i,k)} \in R^{N \times N}$ are weight matrixes in the k th layer,

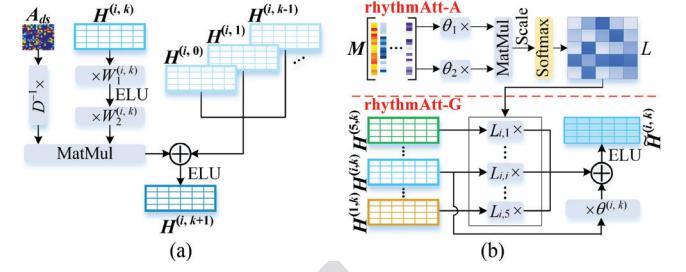


Fig. 5. (a) Structure of the k th resGCN layer on g_i to extract intrarhythm spatiotemporal properties. (b) Illustration of the k th rhythmAtt layer to explore interrhythm complementarity and recalibrate activation under the i th rhythm.

390 and both $\delta_1(\cdot)$ and $\delta_2(\cdot)$ are ELU nonlinearity. Given that
391 $u_i \in R^{E \times 64}$, N is equal to 64. Thereby, with features on
392 node p in g_i at the k th resGCN layer, that is, the p th row
393 of $H^{(i,k)}$, denoted as $h_p^{(i,k)}$, where $p = 1, 2, \dots, E$, the aggre-
394 gator in the k th resGCN for g_i is defined as $\varphi_{\text{res}}^{(i,k)}(\cdot) =$
395 $(\sum_{q=1}^E A_{p,q} \times \delta_1(h_q^{(i,k)} W_1^{(i,k)}) W_2^{(i,k)}) / \sum_{q=1}^E A_{p,q}$, where $A_{p,q}$ is
396 the (p, q) th entry of A_{ds} . In other words, temporal responses
397 due to different ICs, that is, $\delta_1(h_q^{(i,k)} W_1^{(i,k)}) W_2^{(i,k)}$, are selec-
398 tively integrated under the guidance of functional coupling
399 strengths in $A_{p,q} / \sum_{q=1}^E A_{p,q}$. Therefore, irregular spatiotem-
400 poral responses are well gathered by cascaded $\varphi_{\text{res}}^{(i,k)}$ with
401 $k = 0, 1, \dots, K - 1$. However, since a small-word topol-
402 ogy of human brain has been presented in [31], that is, any
403 brain region may be within a relatively small hop neighbor-
404 hood of one another, spatiotemporal attributes gathered by
405 stacked aggregators only retain global structures but discard
406 local ones and, thus, cannot fully depict arbitrary dynamics
407 in an epileptic cortex. To tackle this issue, aggregator $\varphi_{\text{res}}^{(i,k)}$
408 at the k th layer is followed by a residual updater $\eta_{\text{res}}^{(i,k)}(\cdot) =$
409 $\delta_2(h_{Np}^{(i,k)} + \sum_{j=0}^{k-1} h_p^{(i,j)})$, where $h_{Np}^{(i,k)}$ is the neighboring attributes
410 gathered by φ_{res} . Thus, local spatiotemporal features are
411 reserved layer-by-layer via the identity mapping in $\eta_{\text{res}}^{(i,k)}$. As
412 a result, local-global spatiotemporal dynamics under the i th
413 rhythm are well explored as activation $H^{(i,K)} \in R^{E \times 64}$ after K
414 resGCN units. The structure of the k th resGCN layer based
415 on g_i is given in Fig. 5(a). Finally, $H^{(i,K)} \in R^{E \times 64}$ is pooled
416 over nodes to generate compact responses $\gamma_i = m_i \times H^{(i,K)}$,
417 where $m_i \in R^{1 \times E}$ is intrarhythm pooling vector whose j th entry
418 denotes the importance of spatiotemporal properties gathered
419 by the j th IC under the i th rhythm.

420 However, intrarhythm spatiotemporal properties extracted in
421 a parallel way neglect the inter-rhythm complementarity and
422 may lead to overfitting performance consequently. Therefore,
423 a generalized graph convolution, called rhythmAtt, is proposed
424 to model the inter-rhythm dependencies, that is, adjacent
425 matrix L of hierarchical graph G , and integrate spatiotempo-
426 ral attributes under different rhythms selectively. Specifically,
427 inspired by the observation that intrarhythm pooling vector
428 $m_i \in R^{1 \times E}$ globally measures ICs interdependencies under
429 the i th rhythm, submodule rhythmAtt-A first infers inter-
430 rhythm correlations by quantifying similarities between m_i
431 with $i = 1, 2, \dots, 5$ via a scaled dot-product attention [32]

$$432 L = \text{softmax} \left(\frac{1}{\sqrt{d}} (\theta_1 \times M)^T \times (\theta_2 \times M) \right) \quad (4)$$

433 where $M = [\tilde{m}_1^T, \tilde{m}_2^T, \dots, \tilde{m}_5^T]$ ($\tilde{m}_i \in R^{1 \times E}$ is normalized m_i),
 434 and $\theta_1 \in R^{d \times E}$ and $\theta_2 \in R^{d \times E}$ are projection matrixes. Choices
 435 of the scaling factor d will be discussed in Section IV-E. Thus,
 436 impacts of the j th rhythm on the i th rhythm are embedded
 437 in the (i, j) th entry of L . Next, given spatiotemporal activa-
 438 tion under the i th rhythm $H^{(i,k)} \in R^{E \times N}$, namely, features
 439 on node i of hierarchical graph \mathcal{G} at the k th rhythmAtt layer,
 440 where $k = 0, 1, \dots, K$, and submodule rhythmAtt-G encodes
 441 inter-rhythm contextual distribution into intrarhythm $H^{(i,k)}$ to
 442 boost its robustness in a way similar with convolving on the
 443 hierarchical graph \mathcal{G}

$$444 \quad \tilde{H}^{(i,k)} = \delta \left(\sum_{j=1}^5 L_{i,j} H^{(j,k)} + H^{(i,k)} \theta^{(i,k)} \right) \quad (5)$$

445 where $L_{i,j}$ is the (i, j) th entry in L , $\theta^{(i,k)} \in R^{N \times N}$ is the
 446 weight matrix, and $\delta(\cdot)$ is the ELU nonlinearity. Thereby,
 447 aggregator at the k th rhythmAtt-G is factorized as $\varphi_H^{(k)}(\cdot) =$
 $448 H^{(i,k)}(L_{i,i} \times I + \theta^{(i,k)}) + \sum_{j \neq i} L_{i,j} \times H^{(j,k)}$ while the correspond-
 449 ing updater is defined as $\eta_H^{(k)}(\cdot) = \delta(H_{\mathcal{N}_i}^{(i,k)})$, where I is a unit
 450 array, $H_{\mathcal{N}_i}^{(i,k)}$ is the inter-rhythm context gathered by $\varphi_H^{(k)}$. Thus,
 451 resulting $\tilde{H}^{(i,k)}$ not only characterizes spatiotemporal proper-
 452 ties of an epileptic cortex under the specific i th rhythm but also
 453 embeds global distribution of rhythmic activities that are of
 454 other frequency bands, thus enhancing the inter-rhythm com-
 455 plementarity and robustness consequently. A diagram of the
 456 k th rhythmAtt layer to recalibrate the i th rhythm activation
 457 $H^{(i,k)}$ is displayed in Fig. 5(b). Note that all rhythmAtt units in
 458 the HGNC share the same submodule rhythmAtt-A to ensure
 459 a consistent estimation of inter-rhythm dependencies. Finally,
 460 the HGNC maps multirhythm spatiotemporal properties to the
 461 final recognition results via successive fully connected layers.

462 B. Active Preictal Interval Learning

463 With the network defined in the previous section, our
 464 proposed STS-HGCN-AL scheme can theoretically recognize
 465 preictal EEGs from the interictal ones efficiently, after optimiz-
 466 ing its learnable parameters on the training dataset iteratively.
 467 Thus, a high-quality training dataset with correctly annotated
 468 training instances is crucial for obtaining an optimal STS-
 469 HGCN-AL [6]. However, since the preictal transition may
 470 vary from seconds to hours prior to a seizure onset among
 471 different patients, it is intractable to access an optimal preictal
 472 interval empirically and manually annotate training samples
 473 correctly. Thus, the STS-HGCN-AL proposes to infer an
 474 adaptive preictal interval via a semisupervised active learning
 475 approach. A pseudocode of the active preictal interval learning
 476 is shown in Algorithm 1. Specifically, instead of annotating all
 477 training samples at once, active learning strategies only label
 478 a small subset of them at first to initialize model optimization
 479 whereas remaining unlabeled ones will be labeled based on
 480 the updated model outcome in an iterative manner [33]. Thus,
 481 inspired by the observation that recent studies mostly prede-
 482 fine a preictal interval varying from 15 to 60 min [2], [7],
 483 we initialize the minimal preictal interval as 15-min prior to
 484 a seizure onset and the maximum preictal duration as 90-
 485 min before onset of a seizure, respectively. Based on this

Algorithm 1: Semi-Supervised Active Learning Strategy
for the Patient-Specific Optimal Preictal Interval

Input: Initial training dataset $D_t^{(0)}$, initial pooling dataset $D_p^{(0)}$, adaptive preictal interval $\tau_i^{(0)}$ with $i = 1, 2, \dots, N$, confidence threshold λ , uncertainty threshold ς ;
Output: Patient-specific optimal preictal interval τ^* ;
 1 Initializing parameters in the STS-HGCN-AL as $\Theta^{(0)}$;
 2 $q = 0$;
 3 **while** $D_p^{(q)} \neq \emptyset$ **do**
 4 Training $\Theta^{(q)}$ on $D_t^{(q)}$ and updating as $\Theta^{(q+1)}$;
 5 Defining pseudo-labeled pooling subset $D_{pp}^{(q)}$;
 6 Calculating ACE $_{k,i}$ for every $X_{k,i}$ in $D_{pp}^{(q)}$ based on Eq. (6);
 7 Calculating ACM $_{m,i}$ based on Eq. (7);
 8 Defining actual-annotated pooling subset $D_{pa}^{(q)}$;
 9 Updating training dataset:
 10 $D_t^{(q+1)} \leftarrow D_t^{(q)} \cup \{X_{k,i} | X_{k,i} \in D_{pa}^{(q)}, EP(X_{k,i}) > \varsigma\}$;
 11 Updating pooling dataset: $D_p^{(q+1)} \leftarrow D_p^{(q)} \setminus D_{pa}^{(q)}$;
 12 Updating adaptive preictal interval:
 13 $\tau_i^{(q+1)} \leftarrow \tau_i^{(q)} + \underset{m \leq 5}{\operatorname{argmax}}(\text{acM}_{m,i} > \lambda)$,
 $i = 1, 2, \dots, N$;
 14 $q++$
 15 **end**
 16 $\tau = \frac{1}{N} \sum_{i=1}^N \tau_i^{(Q)}, Q = q$.

initialization, given consecutive EEG epochs prior to a spe-
 486 cific seizure, samples within the minimal preictal interval are
 487 assigned to the preictal class while samples out of the maxi-
 488 mum preictal interval are labeled as interictal class. Moreover,
 489 samples that fall in the period between the minimal preictal
 490 interval and maximum preictal period are initially unlabeled
 491 and called poolictal EEGs in the following. As a result, sup-
 492 pose there are total N seizures belonging to a specific patient
 493 used for training, the initial training dataset is obtained as
 494 $D_t^{(0)} = \bigcup_{i=1}^N D_{t,i}^{(0)}$, where $D_{t,i}^{(0)} = \{(X_j, y_j), j = 1, 2, \dots, B_i^{(0)}\}$
 495 is the initial training subset related to the i th seizure with total
 496 $B(0)$ labeled EEG epochs, X_j is a t -second EEG segment, and
 497 $y_j \in \{0, 1\}$ is the corresponding label with 0 and 1 denoting
 498 interictal and preictal classes, respectively. Meanwhile, the ini-
 499 tial pooling dataset is constructed as $D_p^{(0)} = \{D_{p,1}^{(0)}, \dots, D_{p,N}^{(0)}\}$,
 500 where $D_{p,i}^{(0)} = \{X_{k,i}, k = 1, 2, \dots, C_i^{(0)}\}$ consists of total $C_i^{(0)}$
 501 poolictal EEGs related to the i th seizure, and $X_{k,i}$ is the k th
 502 t -second poolictal epoch nearest to the i th seizure onset. Thus,
 503 at the q th training iteration, where $q \geq 0$, parameters of the
 504 STS-HGCN-AL, which is randomly initialized as $\Theta^{(0)}$, are
 505 optimized on the current training dataset $D_t^{(q)}$, and updated as
 506 $\Theta^{(q+1)}$. Next, annotating for samples in the current pooling
 507 dataset $D_p^{(q)}$ can be conducted automatically based on $\Theta^{(q+1)}$,
 508 which requires no extra manual annotation cost. Specifically,
 509 since annotations on EEGs in $D_p^{(q)}$ must be smooth over time,
 510 STS-HGCN-AL first assigns pseudolabels to a subset of $D_p^{(q)}$
 511 based on the temporal continuity, which consists of EEGs
 512

within a 5-min period nearest or furthest to the related seizure onset. Thus, the pseudolabeled pooling subset is denoted as $D_{pp}^{(q)} = \{D_{pp,1}^{(q)}, \dots, D_{pp,N}^{(q)}\}$ with $D_{pp,i}^{(q)} = \{(X_{k,i}, y_{k,i})|X_{k,i} \in D_p^{(q)}, k = 1, 2, \dots, 5M\}$, and $C_i^{(q)}, C_i^{(q)} - 1, C_i^{(q)} - 5M + 1\}$, where $X_{k,i}$ is the k th EEG epoch in $D_{p,i}^{(q)}$, $C_i^{(q)}$ is the total number of EEGs in $D_{p,i}^{(q)}$, $M = 60 \times 5/t$ is the samplewise 1-min resolution (overlapping between temporally adjacent samples is not considered), pseudolabel $y_{k,i}$ is equal to 1, that is, preictal class, if $k \in \{k|1 \leq k \leq 5M\}$, and is equal to 0, that is, interictal class, if $k \in \{k|C_i^{(q)} \geq k \geq C_i^{(q)} - 5M + 1\}$. Next, STS-HGCN-AL defines annotation confidence for $X_{k,i}$ in the $D_{pp}^{(q)}$ by

$$\text{ACE}_{k,i} = \begin{cases} -\sum_{j=1}^2 p(l_j|X_{k,i}, \Theta^{(q+1)}) \\ \log_2(p(l_j|X_{k,i}, \Theta^{(q+1)})) \\ p(y_{k,i}|X_{k,i}, \Theta^{(q+1)}) < 0.5 \\ 2 + \sum_{j=1}^2 p(l_j|X_{k,i}, \Theta^{(q+1)}) \\ \log_2(p(l_j|X_{k,i}, \Theta^{(q+1)})) \\ p(y_{k,i}|X_{k,i}, \Theta^{(q+1)}) \geq 0.5 \end{cases} \quad (6)$$

where $l_j \in \{0, 1\}$ ranges over possible classes, $p(l_j|X_{k,i}, \Theta^{(q+1)})$ is the conditional possibility from current $\Theta^{(q+1)}$ that $X_{k,i}$ belongs to class l_j . Since $p(y_{k,i}|X_{k,i}, \Theta^{(q+1)}) \in [0, 1]$ measures the confidence of $\Theta^{(q+1)}$ on pseudolabel $y_{k,i}$ directly, and $\text{ACE}_{k,i} \in [0, 2]$, which is monotonically increased at $p(y_{k,i}|X_{k,i}, \Theta^{(q+1)})$, is able to quantify annotation confidence on $X_{k,i}$. However, slightly disavowed pseudolabel $y_{k,i}$ by $\Theta^{(q+1)}$, that is, $p(y_{k,i}|X_{k,i}, \Theta^{(q+1)}) \rightarrow 0.5^-$, may be caused by not only an improper $y_{k,i}$ but also currently insufficient recognition capabilities of $\Theta^{(q+1)}$. To alleviate this factor, $\text{ACE}_{k,i}$ rescales annotation confidence drop rate based on the uncertainty degree of $\Theta^{(q+1)}$, that is, given $p(y_{k,i}|X_{k,i}, \Theta^{(q+1)}) \rightarrow 0.5^-, \partial \text{ACE}_{k,i} / \partial p(y_{k,i}|X_{k,i}, \Theta^{(q+1)}) = \log_2(1/p(y_{k,i}|X_{k,i}, \Theta^{(q+1)})) - 1 \rightarrow 0$, which enhances the annotation robustness consequently. However, annotating relying on discrete samples may result in outlier selection. Thus, the proposed STS-HGCN-AL scheme further considers dependence between temporally adjacent samples by averaging adjacent samplewise annotation confidence $\text{ACE}_{k,i}$ within a 1-min interval as $\text{ACM}_{m,i}$,

$$\text{ACM}_{m,i} = \frac{1}{M} \sum_{t=1}^M \text{ACE}_{(m-1) \times M + t, i} \quad (7)$$

where $m = 1, 2, \dots, 10$, $M = 60 \times 5/t$ is the samplewise 1-min resolution. As a result, $X_{k,i}$ in the pseudolabeled pooling subset $D_{pp}^{(q)}$, whose $\text{ACM}_{m,i}$ is upper than confidence threshold λ , will be actually annotated and its corresponding pseudolabel $y_{k,i}$ becomes its actual annotation. λ is priori set as 1. In order to ensure a temporally continuous annotation, subset of $D_p^{(q)}$ that is actual-labeled is defined as $D_{pa}^{(q)} = \{D_{pa,1}^{(q)}, \dots, D_{pa,N}^{(q)}\}$ with $D_{pa,i}^{(q)} = \{(X_{k,i}, y_{k,i})|X_{k,i} \in D_p^{(q)}, k = 1, \dots, \mu_{i,1}M, C_i^{(q)}, \dots, C_i^{(q)} - \mu_{i,0}M + 1\}$, where $\mu_{i,1} = \text{argmax}_{m \leq 5} (\text{ACM}_{m,i} > \lambda)$ and $\mu_{i,0} = \text{argmax}_{m > 5} (\text{ACM}_{m,i} > \lambda) - 5$, $X_{k,i}$ is the k th segment in $D_{p,i}^{(q)}$, $M = 60 \times 5/t$ is the samplewise 1-min resolution, and $C_i^{(q)}$ is the total number of epochs

in $D_{p,i}^{(q)}$. In other words, poolictal EEGs within a $\mu_{i,1}$ -minute interval nearest to the related seizure onset are gathered to the preictal class while poolictal EEGs within a $\mu_{i,0}$ -minute interval furthest to the related onset are merged to interictal class. Finally, the pooling dataset $D_p^{(q)}$ is updated as $D_p^{(q+1)}$ by removing $D_{pa}^{(q)}$. The training dataset $D_t^{(q)}$ is updated as $D_t^{(q+1)}$ by selectively appending informative instances in the $D_{pa}^{(q)}$. The informativeness of samples in the $D_{pa}^{(q)}$ is quantified by using the Shannon entropy (EP), since larger EP, that is, upper than an uncertainty threshold ζ , indicates that corresponding sample is uncertainly identified by the current $\Theta^{(q+1)}$ and is more worth to be trained on [34]. ζ is empirically set as 0.9. Adaptive preictal interval $\tau_i^{(q)}$ related to the i th seizure is updated as $\tau_i^{(q+1)} = \tau_i^{(q)} + \mu_{i,1}$, where $\tau_i^{(0)} = 15$ (unit: min), $i = 1, 2, \dots, N$. As a result, after total Q round active preictal interval inferring, every poolictal epoch is assigned to the preictal or interictal class properly, and the patient-specific optimal preictal interval τ^* is obtained as $\tau^* = (\sum_i^N 1\tau_i^{(Q)})/N$. Model optimization will continue on the final training dataset $D_t^{(Q)}$ until training stop criterion is satisfied and, thus, an optimal STS-HGCN-AL scheme for seizure prediction is obtained.

C. Postprocessing and Implement Details

In this section, we translate obtained optimal STS-GCN-AL scheme into a practical seizure predictor via a persistent warning scheme [35]. Specifically, given consecutive EEG epochs, a conditional probability series $p(i)$ is first obtained, whose i th entity denotes the conditional probability from the STS-GCN-AL method that the i th epoch belongs to preictal class. $p(i)$ is then smoothed over time by a moving average filter to filter out outliers and obtain the reliable probability signal $p_s(i)$. Inspired by latest seizure prediction studies [2], the length of the moving average filter is set as 1 min, whose influence over seizure prediction performance will be discussed in Section IV-G. Thus, when $p_s(i)$ exceeds a threshold ω ($\omega = 0.6$) at t_a , a timer $s(i)$ of duration τ_ω is triggered to warn the patient for a seizure imminent [35]. Although ω is priori determined based on the existing studies [6], its influence over seizure warning performance will be discussed in Section IV-G. τ_ω indicates persistence parameter and is essentially equal to the patient-specific optimal preictal interval τ^* in the previous section [35]. For a correct warning, $s(i)$ must start at a t_a that is at least $\tau_{\omega 0}$ prior to the seizure onset, which is referred to as the detection interval [35], and remain activated until the seizure starts. Otherwise, it is a false alarm. Although a long-duration $\tau_{\omega 0}$ enables the patient to be more prepared for the upcoming seizure, recent studies merely define a $\tau_{\omega 0}$ less than 1 min, due to the severe uncertainty of seizures [2], [35]. In order to provide a $\tau_{\omega 0}$ as long as possible under the premise of a correct warning, we infer an incremental $\tau_{\omega 0}$ as $\tau_{\omega 0} = \tau/\tau^{(0)}$, where τ^* and $\tau^{(0)}$ are the patient-specific optimal preictal interval and the minimal preictal interval (15 min) in the previous section. As a result, an adaptive seizure warning system, which considers interactions with patients carefully [35], is produced.

TABLE I
DATA INFORMATION OF CHB-MIT DATABASE

Patient ID	No. of seizures	Total EEGs in hour	Patient ID	No. of seizures	Total EEGs in hour
Age	Gender		Age	Gender	
1	11	F	7	34.6	
2	11	M	3	27.6	
3	14	F	6	28.9	
5	7	F	5	19.1	
6	1.5	F	7	30.9	
8	3.5	M	5	11.5	
9	10	F	4	54.3	
10	3	M	6	29.3	
11	12	F	3	32.5	
13	3	F	7	18.3	

where M and F represent male and female patients respectively.

In detail, our STS-HGCN-AL network is built on the Pytorch framework [16]. The initial parameters in the STS-HGCN-AL, that is, $\Theta^{(0)}$ in the previous section, are generated by the default Pytorch Kaiming weight initializer. The convergence of parameters in the STS-HGCN-AL during training is achieved by optimizing the cross-entropy loss function via an Adam optimizer with a batch size of 128 and a learning rate of 0.001. Training will be stopped when accuracies on the training dataset no longer increase for consecutive 15 iterations.

IV. EXPERIMENTAL RESULTS

A. EEG Database and Experimental Settings

Experiments are carried out on the CHB-MIT Scalp EEG Database [36], which consists of long-term EEG recordings at a sampling rate of 256 Hz from 23 patients with intractable seizures. The details about the CHB-MIT database are given in Table I. In this study, patients with at least a 3-h interictal period and at least two seizures are included for the evaluation of seizure prediction performance [2]. In order to exclude the effect of the postictal period, recordings within 2-h after a seizure are removed. Moreover, if several seizures cluster within a 2-h period, only the leading one is retained for the evaluation of seizure prediction performance [35]. Continuous EEG recordings are divided into nonoverlapping 5-s epochs before being fed into the STS-HGCN-AL scheme.

In order to obtain an unbiased evaluation of the seizure prediction performance, patient-specific leave-one-out (LOO) cross-validation is used. Specifically, for the i th patient, suppose there are total N seizures in the database, in each LOO loop, one seizure is left for testing while other ($N-1$) seizures are used for training. This procedure repeats N times until all seizures in the i th patient have been held out for evaluation. Reported metrics for the i th patient are an average across N loops. The area under curve (AUC) is used to evaluate the classification performance of our STS-HGCN-AL scheme. In order to ensure a reliable comparison with nonactive baseline methods that could only utilize training EEGs in the initial training dataset defined in Section III-B for model optimization, AUC is evaluated on a testing dataset defined based on the same criteria as that of the initial training dataset. Sensitivity (S_n) and false-positive rate (FDR/h), which are defined as the ratio of correctly predicted seizures to the total number of seizures and the number of false alarms per

hour, respectively, are used to assess our scheme as a seizure warning system. For nonactive baseline methods, persistence parameter τ_ω is set as 15 min, that is, the initial status in our experiment. Correspondingly, detection interval τ_{ω_0} is set as 1 min. Moreover, the significance of an improvement over chance-level (p -value) is evaluated to understand whether our seizure warning system is statistically better than a random predictor. For a detailed mathematical derivation, refer to [35].

B. Overall Performance

Patient-specific overall performance of the STS-HCN-AL framework is evaluated among approaches as follows.

- 1) Empirical mode decomposition (EMD)-based temporal and spectral analysis with a back-end support vector machine classifier (SVM) is a commonly used baseline method to recognize epileptic EEGs and predict seizures [37].
- 2) Deep ConvNet utilizes regular convolutions to exploit both spatial and temporal EEG dependencies. It is currently one of the most popular deep learning approaches designed for EEG decoding [30].
- 3) DCNN+Bi-LSTM uses a deep convolutional network to extract EEG spatial couplings and a back-end bidirectional long short-term memory to exploit EEG temporal responses. It is a state-of-the-art deep learning strategy that is typically designed for EEG seizure prediction [7].

From Table II, the proposed STS-HGCN-AL scheme yields an average AUC of 0.938 while other baseline methods only achieve an average AUC of 0.791, 0.874, and 0.824, respectively, indicating the robust classification abilities of our proposed method with respect to the changes of discrimination threshold. Especially, patients 1, 8, 18, 22, and 23 reach an AUC greater than 0.99, which proves the efficacy of our method in distinguishing preictal EEGs from interictal ones. Moreover, in the seizure prediction scenario, our seizure predictor outperforms all the compared studies by successfully warning a total of 93 out of 98 seizures. Meanwhile, our system yields an average FDR/h of 0.109/h, which is at least 39.1% improvement against baseline methods. As a result, the proposed method achieves a promising S_n with a negligible FDR/h and validates its abilities in predicting seizures intuitively. In addition, the improvement-over-chance of our EEG seizure predictor is statistically significant under 99% confidence interval ($p < 0.01$) for a total of 17 out of 19 patients, and the p -values of the proposed seizure warning system versus the random one are less than 0.05 for all patients, proving the significantly excellent seizure prediction performance of the proposed STS-HGCN-AL framework.

C. Impact of the ST-SENet

Based on the above overall performance, our STS-HGCN-AL scheme can predict seizures effectively in the premise of estimating temporal evolutions in an epileptic cortex with the ST-SENet. We further exploit the efficacy of the ST-SENet in capturing temporal cortical dependencies by comparing the STS-HGCN-AL with ST-SENet against a counterpart that

TABLE II
PATIENT-SPECIFIC OVERALL COMPARISON OF PERFORMANCE ON CHB-MIT DATABASE

patient ID	EMD+SVM [37]				Deep ConvNet [30]				DCNN+Bi-LSTM [7]				Proposed STS-HGCN-AL			
	AUC	S_n (%)	FDR/h	p-value	AUC	S_n (%)	FDR/h	p-value	AUC	S_n (%)	FDR/h	p-value	AUC	S_n (%)	FDR/h	p-value
1	0.992	100.0	0.029	<0.001	0.995	100.0	0.000	<0.001	0.944	100.0	0.116	0.005	0.996	100.0	0.000	<0.001
2	0.666	66.7	0.181	0.337	0.605	100.0	0.145	0.124	0.709	66.7	0.254	0.376	0.897	100.0	0.145	0.042
3	0.638	83.3	0.173	<0.001	0.906	83.3	0.277	0.001	0.738	66.7	0.277	0.042	0.928	83.3	0.173	0.001
5	0.804	100.0	0.209	0.011	0.804	80.0	0.262	0.075	0.846	80.0	0.157	0.013	0.875	100.0	0.000	0.001
6	0.723	85.7	0.259	0.062	0.708	85.7	0.356	0.006	0.779	85.7	0.356	0.004	0.906	100.0	0.162	<0.001
8	0.997	100.0	0.087	0.002	0.999	100.0	0.000	0.001	0.982	100.0	0.174	0.002	0.999	100.0	0.000	0.001
9	0.481	100.0	0.645	0.029	0.653	50.0	0.331	0.388	0.717	75.0	0.424	0.234	0.843	100.0	0.092	<0.001
10	0.805	83.3	0.478	0.002	0.935	83.3	0.410	<0.001	0.866	66.7	0.478	0.023	0.977	83.3	0.171	<0.001
11	0.935	100.0	0.308	0.002	0.932	100.0	0.123	<0.001	0.884	100.0	0.400	0.009	0.940	100.0	0.123	<0.001
13	0.874	85.7	0.328	0.011	0.871	85.7	0.219	0.005	0.786	85.7	0.219	0.001	0.915	85.7	0.109	<0.001
14	0.766	100.0	0.417	0.070	0.971	100.0	0.104	<0.001	0.710	100.0	0.313	0.031	0.976	100.0	0.104	<0.001
16	0.868	87.5	0.561	0.022	0.954	87.5	0.280	<0.001	0.758	87.5	0.374	0.009	0.954	87.5	0.187	<0.001
17	0.561	33.3	0.237	0.270	0.732	66.7	0.237	0.208	0.822	66.7	0.355	0.149	0.826	100.0	0.237	0.007
18	0.742	75.0	0.277	<0.001	0.953	75.0	0.138	0.001	0.955	50.0	0.208	0.036	0.992	75.0	0.138	<0.001
19	0.794	100.0	0.038	<0.001	0.803	0.0	0.038	0.400	0.990	0.0	0.038	0.302	0.991	100.0	0.038	0.001
20	0.944	100.0	0.184	<0.001	0.982	100.0	0.184	<0.001	0.960	100.0	0.184	<0.001	0.982	100.0	0.184	<0.001
21	0.744	100.0	0.313	0.007	0.821	100.0	0.208	0.008	0.648	100.0	0.521	0.101	0.833	100.0	0.156	0.017
22	0.733	100.0	0.435	0.024	0.997	100.0	0.000	0.001	0.720	100.0	0.261	0.044	0.997	100.0	0.000	0.001
23	0.954	100.0	0.284	<0.001	0.989	100.0	0.095	<0.001	0.839	85.7	0.095	<0.001	0.990	100.0	0.047	<0.001
Aver	0.791	89.5	0.287	—	0.874	84.1	0.179	—	0.824	79.8	0.274	—	0.938	95.5	0.109	—

where significance levels of >0.05 are marked with bolds.

could only use a handcrafted feature set to characterize temporal cortical evolutions. Specifically, we extract spectral power, mean, variance, skewness, kurtosis, mobility, and complexity statistics of the Hjorth parameter from δ , θ , α , β , and γ rhythms, respectively, to characterize the temporal dependencies in an epileptic brain [2] and serve as the node attributes in the hierarchical graph \mathcal{G} . From Fig. 6, it is observed that AUCs of all patients have been boosted. Especially, a maximum AUC increasing of 0.306 is reached by patient 9, which proves the effectiveness of the ST-SENet in capturing critical temporal responses under the preictal transition. Moreover, a significant drop of the standard deviation of AUCs across different patients is achieved from 0.109 to 0.060, which shows that inferring EEG graph representation in a data-driven manner can combat the heterogeneity among patients efficiently. These results prove that the ST-SENet can indeed explore the temporal cortical evolutions and mitigate the variation of seizures efficiently.

In addition, we evaluate the impacts of each one of five rhythms on the classification performance of our STS-HGCN-AL, and results are given in Table III. As for the single rhythm, the average AUC varies from 0.828 (α rhythm) to 0.914 (δ rhythm), which indicates that neuronal activities with respect to different frequency bands include different information for the seizure anticipation and emphasizes the necessity of multiresolution spectral analysis for seizure prediction intuitively. It should be noted that the average AUC using only δ rhythm reaches 0.914, while the maximum AUC for patient 8 from θ rhythm solely achieves 0.998. This may be because that abnormal seizure waves may occur at different frequency rhythms but are mainly distributed in the range of 3–6 Hz [38].

D. Influence of the GATENet

Apart from temporal exploration in the ST-SENet, the efficacy of the GATENet in inferring ICs spatial couplings also affects the representational abilities of the resulting hierarchical graph \mathcal{G} . To assess the efficacy of the GATENet in

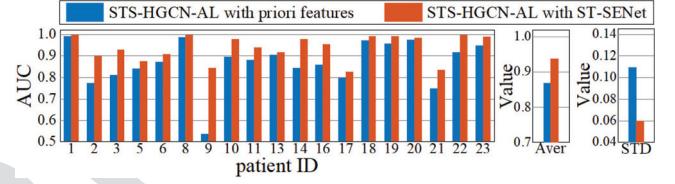


Fig. 6. UC comparison with and without ST-SENet.

TABLE III
AUC COMPARISON WITH RESPECT TO DIFFERENT RHYTHMS

ID	δ rhythm	θ rhythm	α rhythm	β rhythm	γ rhythm	Overall
1	0.998	0.996	0.813	0.975	0.97	0.996
2	0.677	0.656	0.806	0.682	0.761	0.897
3	0.883	0.88	0.812	0.852	0.836	0.928
5	0.892	0.873	0.843	0.817	0.843	0.875
6	0.882	0.873	0.825	0.822	0.886	0.906
8	0.999	0.998	0.75	0.979	0.979	0.999
9	0.659	0.641	0.834	0.686	0.662	0.843
10	0.956	0.951	0.873	0.933	0.948	0.977
11	0.912	0.891	0.682	0.871	0.884	0.94
13	0.923	0.922	0.883	0.906	0.899	0.915
14	0.982	0.983	0.795	0.906	0.929	0.976
16	0.954	0.901	0.832	0.887	0.849	0.954
17	0.858	0.816	0.803	0.778	0.765	0.826
18	0.986	0.716	0.833	0.745	0.693	0.992
19	0.99	0.991	0.937	0.973	0.975	0.991
20	0.993	0.997	0.852	0.956	0.946	0.982
21	0.832	0.821	0.848	0.726	0.717	0.833
22	0.996	0.996	0.825	0.976	0.977	0.997
23	0.987	0.966	0.879	0.967	0.969	0.99
Aver	0.914	0.888	0.828	0.865	0.868	0.938

where bold font indicates the best results.

capturing critical epileptic interactions, the GATENet is compared with two counterparts and experimental results are given in Fig. 7(a).

- 1) STS-HGCN-AL with random is constructed by replacing the GATENet with the Eedös-Rényi random graph with an edge probability of 0.5 [12]. Therefore, no ICs spatial correlations can be learned.
- 2) STS-HGCN-AL with PLV is obtained by estimating ICs spatial couplings with PLV rather than GATENet, where PLV is a manual static that infers functional connection strengths between EEGs as the absolute value

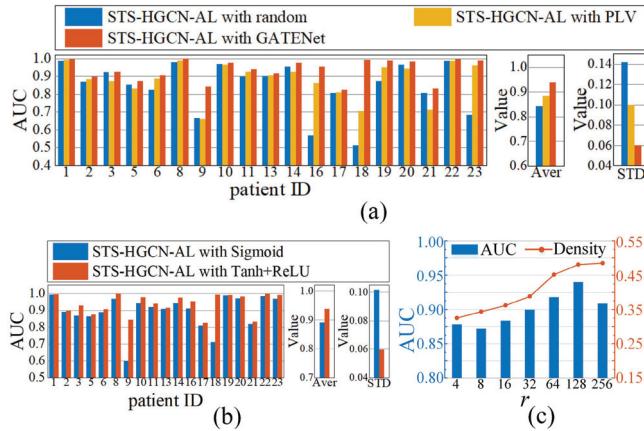


Fig. 7. (a) AUC comparison with respect to GATENet. (b) AUC comparison depending on sparsity. (c) Performance with different r .

of the mean phase difference between the two signals expressed as a complex unit-length vector [13]. Thus, the STS-HGCN-AL with PLV can only use the prior indicator to infer ICs spatial correlations and may ignore the diversity among patients.

From Fig. 7(a), STS-HGCN-AL with PLV gains an average 0.039 higher AUC than the counterpart with a randomly initialized topology by taking ICs spatial coupling strengths into account, which shows the necessity of inferring functional correlations between brain regions in characterizing the preictal transition. However, it is not sufficient to merely use manual indicators to quantify the intractable interactions in an epileptic brain, due to the severe variation of seizures. Therefore, by inferring ICs functional connectivity in a data-driven way, the proposed STS-HGCN-AL with GATENet further gains higher patient-specific AUC than the counterpart with PLV ranging from 0.003 to 0.286. These results prove that the GATENet can indeed capture abnormal interactions in an epileptic brain and boost the robustness of the resulting graphs in representing preictal EEGs.

In order to prune the spurious interactions between IC pairs, the GATENet employs a combination of Tanh and ReLU nonlinearities to penalty weak ICs couplings and result in a sparse brain network. To validate the necessity of sparse regularization, we compare the STS-HGCN-AL with Tanh+ReLU against a counterpart with a Sigmoid replacement. Thus, the STS-HGCN-AL with Sigmoid does not prune weak ICs correlations and produces a dense cortex topology. From Fig. 7(b), a consistent AUC increasing ranging from 0.003 to 0.282 is observed across all patients, proving that sparsity constraint is beneficial for detecting an optimal topology of the epileptic brain. In addition, the reduction ratio r also controls the degree of sparsity in the resulting A_{ds} . We set r varying from {4, 8, 16, 32, 64, 128, 256} to understand its influence on the recognition performance, since the patient-specific number of EEG electrodes in the CHB-MIT database is usually less than 23. From Fig. 7(c), density of the A_{ds} , which is defined as the fraction of present connections to possible connections in A_{ds} , grows with increasing r . Thus, a large r detects a dense yet potentially

TABLE IV
PERFORMANCE COMPARISON WITH RESPECT TO HGCN

ID	MLP		GCN		resGCN		resGCN+rhythmAtt	
	AUC	p-value	AUC	p-value	AUC	p-value	AUC	p-value
1	0.993	<0.001	0.995	<0.001	0.991	<0.001	0.996	<0.001
2	0.688	0.082	0.810	0.114	0.868	0.082	0.897	0.042
3	0.889	0.001	0.865	0.002	0.899	0.001	0.928	0.001
5	0.801	0.005	0.872	0.001	0.872	0.001	0.875	0.001
6	0.864	0.004	0.878	0.004	0.904	0.002	0.906	<0.001
8	0.997	0.001	0.998	0.001	0.999	0.001	0.999	0.001
9	0.581	0.299	0.650	0.327	0.683	0.283	0.843	<0.001
10	0.978	<0.001	0.971	<0.001	0.976	0.006	0.977	<0.001
11	0.905	<0.001	0.906	<0.001	0.939	<0.001	0.940	<0.001
13	0.900	<0.001	0.907	<0.001	0.911	0.001	0.915	<0.001
14	0.930	<0.001	0.966	<0.001	0.975	<0.001	0.976	<0.001
16	0.896	0.002	0.919	<0.001	0.939	<0.001	0.954	<0.001
17	0.818	0.113	0.816	0.010	0.816	0.009	0.826	0.007
18	0.688	0.007	0.930	<0.001	0.984	<0.001	0.992	<0.001
19	0.944	0.117	0.963	0.325	0.964	0.002	0.991	0.001
20	0.975	<0.001	0.973	<0.001	0.980	<0.001	0.982	<0.001
21	0.749	0.012	0.806	0.020	0.826	0.010	0.833	0.017
22	0.992	0.016	0.996	0.001	0.996	0.007	0.997	0.001
23	0.984	<0.001	0.985	<0.001	0.986	<0.001	0.990	<0.001
Aver	0.872	—	0.906	—	0.922	—	0.938	—

where significance levels of >0.05 are marked with bolds.

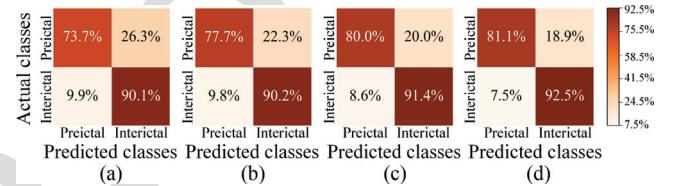


Fig. 8. Cross-patient confusion matrices. (a) MLP. (b) GCN. (c) resGCN. (d) resGCN+rhythmAtt.

spurious cortex topology while a small one results in a possibly over-shrinking one. As a result, the best average AUC across different patients is achieved by a moderate $r = 128$.

E. Classification Performance Depending on the HGCN

In order to reconcile multirhythm spatiotemporal dynamics in epileptic activities, resGCN and rhythmAtt mechanisms in the subnet HGCN are adopted to capture and integrate intrarhythm spatiotemporal dependencies, respectively. We further compare the HGCN among baseline approaches as follows to prove its efficacy and results are given in Table IV.

- 1) MLP is obtained by removing both resGCN and rhythmAtt units in the HGCN. Thus, MLP can only explore temporal-spectral evolutions while spatial ones cannot be learned.
- 2) GCN is implemented by replacing resGCN and rhythmAtt units with a general form of graph convolutions in [39]. Thus, although global spatio-temporal-spectral properties can be exploited by GCN, both local structures and inter-rhythm supplementary cannot be considered.
- 3) resGCN is constructed by removing rhythmAtt units in the HGCN. Thus, although both global and local intrarhythm spatiotemporal evolutions in an epileptic cortex can be exploited, inter-rhythm complementarity is not learned.

From Table IV, the combination of resGCN and rhythmAtt units outperforms all baseline methods at all patients in

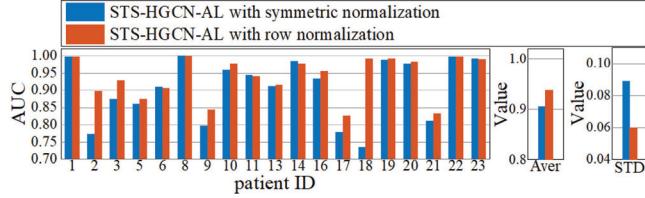


Fig. 9. AUC comparison with respect to row normalization.

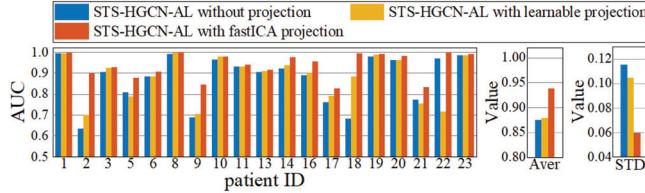


Fig. 10. AUC comparison with respect to fastICA projection.

terms of AUC, proving the efficacy of weighting multirhythm spatiotemporal properties for recognizing preictal EEGs. Specifically, by extracting spatio-temporal-spectral attributes jointly, GCN yields a higher average AUC of 0.034 against MLP. However, the local dynamics in an epileptic cortex are potentially discarded by the GCN. Thus, resGCN gains an AUC improvement of 0.016 by retaining both global and local spatiotemporal cortical dynamics simultaneously. Finally, the proposed HGCN, where resGCN and rhythmAtt units are iteratively used to not only capture spatiotemporal properties under various rhythms but also integrate multirhythm attributes selectively, yields an at most 0.160 increasing of AUC and, thus, predicts seizures more efficiently. Moreover, the improvement over-chance is statistically insignificant for four patients in MLP, three patients in GCN, and two patients in resGCN, while our HGCN is statistically better than chance at all patients, at a significance value of 0.05. In addition, we exploit the efficacy of the HGCN by a cross-patient confusion matrix. Similar performance improvement is observed in Fig. 8.

As for the scale factor d in rhythmAtt unit, we set d varying from {2, 4, 8, 16} to evaluate its impacts on the classification performance, since the EEG channels in CHB-MIT database are commonly less than 23 electrodes. From Table V, the best average AUC across all patients is achieved by $d = 16$. The possible reason is that a small d may result in inadequate learnable parameters to fully model the inter-rhythm complementarity.

Moreover, although we empirically adopt row normalization for the resGCN unit, that is, $D^{-1} \times A_{ds}$ in (3), we still evaluate its efficacy by replacing it with a symmetric normalization, namely $D^{-0.5} \times A_{ds} \times D^{-0.5}$, which is also observed in some GCN studies [10], [15]. From Fig. 9, our STS-HGCN-AL with row normalization is stronger than its counterpart with symmetric normalization with an average AUC improvement of 0.032. This is possibly because although both normalization strategies could prevent the resulting activation from growing too large, row normalization emphasizes directional influence between various brain regions and, thus, owns better performance for seizure prediction.

Apart from the network design, prior fastICA projection also contributes to the promising performance of the HGCN since it ensures an input graph to represent the epileptic cortex directly rather than merely the mixed scalp EEGs. We further compare the STS-HGCN-AL with fastICA against two baselines as follows to validate the efficacy of prior source projection.

- 1) STS-HGCN-AL without projection is implemented by removing prior fastICA projection and, thus, no subsequent analysis in the source domain is allowed.
- 2) STS-HGCN-AL with learnable projection is conducted by replacing the fastICA source projection with a preceded learnable parameter in the network. Therefore, no mutual independence across the resulting series is guaranteed.

From Fig. 10, the STS-HGCN-AL with fastICA outperforms all baselines with a consistent AUC improvement ranging from 0.0003 to 0.310, which validates the efficacy of regularizing mutual independence across node attributes in resulting graphs.

F. Performance by the Active Preictal Interval Learning

In order to combat the diversity of preictal transitions, a semisupervised active learning strategy is nested in our STS-HGCN-AL to infer an optimal preictal interval τ^* in every patient-specific LOO loop. To evaluate the efficacy of our active preictal interval estimation, we first confirm whether it is statistically better than fix ones using the Wilcoxon signed-rank test for AUCs in Fig. 11(a), where the p -value is reported by the vertical axis to the horizontal axis. Since a preictal interval varying from 15 to 60 min was usually defined in [2] and [7], contrast fixed preictal duration in our experiment ranges in {15, 30, 60, 90} (unit: min). From Fig. 11(a), none of the fixed preictal intervals can be statistically better than all others, indicating the importance of learning optimal one adaptively. Thus, actively inferred τ^* in our method yields the best results and is significantly superior to all fixed selections at a level of 0.05, which proves the advantages of our active preictal interval learning strategy intuitively. In addition, the auxiliary training samples, which are selectively added into the initial training set, also contribute to the promising recognition performance of our method. Thus, we then assess the STS-HGCN-AL method trained with or without auxiliary samples by the patient-specific AUCs in Fig. 11(b). Note that due to the EEG recordings discontinuity, no poolictal segment, that is, potentially auxiliary training instances, is included in patients 11, 14, 17, 18, and 19, and the corresponding comparison in Fig. 11(b) is *null*. From Fig. 11(b), AUCs of almost all patients have been boosted, showing that appending informative training samples can indeed boost the robustness of our scheme. In detail, resulting patient-specific optimal preictal intervals are given in Fig. 11(c), with means and standard deviations across LOO loops given. From Fig. 11(c), τ^* varies both intrapatients and interpatients due to the severe variation of seizures, which emphasizes the necessity of inferring adaptive preictal interval instead of predefining fixed ones once more. Moreover, an increased ratio by the baseline preictal interval

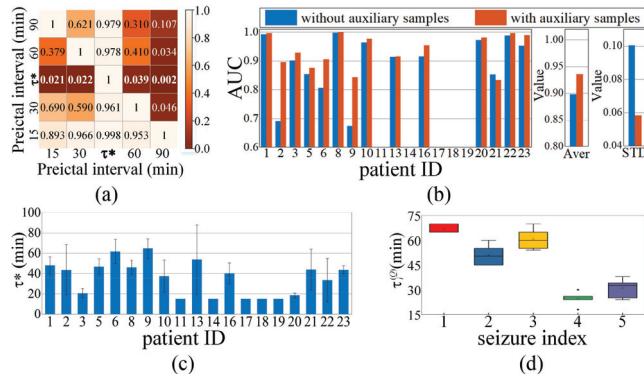


Fig. 11. (a) Significance matrix with respect to the superiority of different preictal intervals. (b) AUC comparison with and without auxiliary training samples. (c) Patient-specific optimal preictal interval τ^* . (d) Convergent preictal interval $\tau_i^{(Q)}$ during training stage for a specific patient.

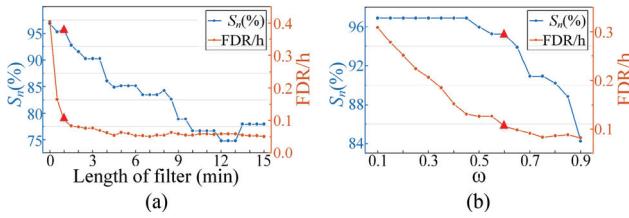


Fig. 12. Performance comparison of the seizure prediction with different parameters. (a) Filter length. (b) Threshold ω .

TABLE V
AUC COMPARISON OF OUR METHOD WITH DIFFERENT SCALE FACTOR d

Scale factor d	Aver AUC \pm STD	Parameters ($\times 10^3$)
2	0.889 ± 0.117	0.86
4	0.932 ± 0.063	1.62
8	0.935 ± 0.058	3.14
16	0.938 ± 0.060	6.18

where bold fonts indicate best results.

(15 min) enables an average 37% increasing of detection interval $\tau_{\omega 0}$, which provides the patient with more sufficient time to prepare for the incoming seizure. Especially, to validate the stability of our active preictal interval learning, we visualize the convergent $\tau_i^{(Q)}$ during the LOO training stage for each seizure of a specific patient in Fig. 11(d). Outlier inferring across LOO loops only exists in the fourth seizure, proving the appealing convergence abilities of our preictal interval learning strategy.

G. Influence of Postprocessing

In order to filter out prediction outliers, the first step in postprocessing is to smooth the model outcome with a moving average filter. We explore the sensitivity of the filter length in the range from 0 to 15 with a step size of 0.5 (unit: min), where 0 min means without moving average filter. The tradeoff between S_n and FDR/h can be seen in Fig. 12(a). The possible reason is that an overly large filter length may result in an oversmoothing probability signal, causing some short-duration warning to be missed, that is, unsatisfying S_n , while for a small filter length, prediction outliers may still exist, resulting in an over-sensitive seizure predictor, that is, poor FDR/h. Therefore,

TABLE VI
COMPARISON OF MODEL COMPACTNESS

Model	Seizure prediction		Model compactness	
	S_n (%)	FDR/h	Parameters ($\times 10^3$)	Inference time (ms)
Deep ConvNet [30]	84.1	0.179	0.282	3.98
DCNN-Bi-LSTM [7]	79.8	0.274	0.023	21.94
CE-stSENet [28]	—	—	0.351	179.50
This work	95.5	0.109	0.341	36.90

where bold fonts indicate our proposed method.

the filter length in this study is set as a moderate 1 min, which is in line with the previous studies [2].

Apart from the filter length, the threshold ω also affects the performance of our seizure warning system, since only the probability signal exceeds ω , the system will declare a seizure imminent. We further vary ω in ranging from 0.1 to 0.9 with a step size of 0.05 to evaluate its sensitivity. The performance tradeoff between S_n and FDR/h is also observed in Fig. 12(b), which is similar to the properties of the filter length on our method. Thus, in order to realize robust seizure prediction, we set ω as a moderate 0.6, which is the same value as the existing studies [6].

V. DISCUSSION

A. Efficacy of Model Compactness

To demonstrate the compactness of our method, we compare the model sizes between our STS-HGCN-AL network and recent methods, and results are exhibited in Table VI, where the CE-stSENet is an EEG seizure detection scheme discussed in our previous study [28]. Although the proposed STS-HGCN-AL, which involves roughly 3.4×10^5 learnable parameters, is much more complex than the DCNN-Bi-LSTM, the complexity of our STS-HGCN-AL is of the same order as that of the Deep ConvNet and CE-stSENet. Since the real-time study is one of the most important factors for an online seizure warning system, we further evaluate the inference time, which is defined as the required time for a deep learning model to make a decision for one EEG epoch, and results are listed in Table VI. We perform this procedure on a GeForce RTX 2080 GPU with 8-GB memory. From Table VI, although our model requires a longer time to process an EEG epoch than baseline methods, our inference time, that is, 36.9 ms, is significantly less than the length of an EEG segment, that is, 5 s, which enables our scheme to be applied in a practical real-time seizure prediction scenario. Moreover, our STS-HGCN-AL can achieve the best seizure prediction performance in terms of S_n and FDR/h without extra computational complexity involved.

B. Efficacy of Predicting Seizures in Time

Based on the performance comparison discussed above, our STS-HGCN-AL scheme can indeed predict seizures efficiently with a promising S_n and a negligible FDR/h. To further validate whether seizures can be warned timely, the proposed STS-HGCN-AL is compared with other seizure predictors in terms of average prediction time (APT), which is defined as the average of intervals between warning timer and the related

TABLE VII
EXPERIMENTAL SETTINGS AND PERFORMANCE RESULTS OF EXISTING METHODS ON CHB-MIT DATABASE

Authors	Methods	No. of patients-seizures	Validation scheme	Interictal-preictal intervals (min)	Evaluated hours	No. of patients over chance	Average AUC- S_n (%) FDR/h
Zandi et.al [45]	Zero-crossing intervals+ Bayesian Gaussian mixture	3-18	no CV	40-40	273	3/3	NR-83.81-0.165
Cho et.al [43]	EMD, PLV+SVM	21-65	10-fold CV	30-5	10.85	NR	NR-82.44-NR
Truong et.al [44]	Short Time Fourier Transform+CNN	13-64	LOOCV	240-30	209	12/13	81.2-0.160
Khan et.al [6]	Wavelet transform+CNN	15-18	10-fold CV	10-10	70.5	NR	0.866-87.8-0.147
Tsiouris et.al [5]	Temporal, frequency, correlation, graph theory analysis+LSTM	24-185	10-fold CV	15-15 30-30 60-60 120-120	979.9	NR NR NR NR	NR-100.0-0.110 NR-100.0-0.060 NR-100.0-0.030 NR-100.0-0.020
Ozcan et.al [2]	Spectral power, statistical moments, Hjorth+3D CNN	16-77	LOOCV	60-60 120-60 240-60	466.1 419.4 353.5	13/16 14/16 15/16	NR-86.8-0.292 NR-87.0-0.186 NR-85.7-0.096
This work	STS-HGCN-AL	19-98	LOOCV	Adaptive	453.0	19/19	0.938-95.5-0.109

where bold fonts are our method, No. of patients-seizures are total number of patients and seizures involved for evaluation, evaluated hours are cross-patient summation of total EEGs in hour in Table I.

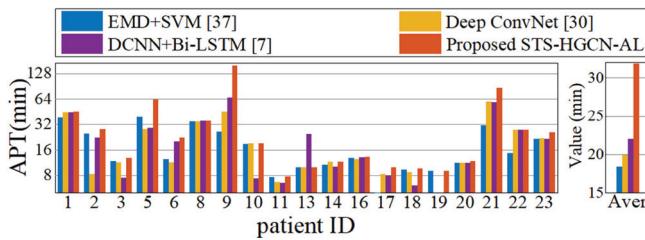


Fig. 13. APT comparison among different seizure predictors.

seizure onset, and experimental results are displayed in Fig. 13. Note that if a seizure is failed to be forecasted, the corresponding APT is equal to 0. From Fig. 13, our STS-HGCN-AL approach yields a cross-patient average APT of 31.8 min, which is at least 10.6% increasing against baseline predictors. Especially, for patient 9, our predictor advances seizure onset with an average APT of 155.1 min, which provides the patient with a long enough preparation for the upcoming seizures efficiently. At a result, our STS-HGCN-AL is capable of not only forecasting seizures reliably with competitive S_n and FDR/h but also warning patients enough ahead of time with a significant APT.

999 C. Efficacy of Pathological Spatio-Temporal-Spectral 1000 Dependencies Exploration

In order to interpret the spectral-temporal evolutions in an epileptic brain captured by our novel STS-HGCN-AL scheme, we visualize feature maps learned by the last gcSE unit in the ST-SENet via a deconvolution scheme [40] and results for a 2-h duration prior to a specific seizure are shown in Fig. 14(a). From Fig. 14(a), our predictor successfully warns the seizure imminent roughly 65 min ahead of the seizure onset. Compared with interictal control recordings, which exhibit no significant inter-rhythm energy shifting, preictal EEGs tend to activate the STS-HGCN-AL, especially in θ and δ rhythms. This observation is agreed to previous studies that seizure propagation may be characterized by a shifting from fast rhythm activities in spatially focal regions to slower frequencies across widespread areas [17]. Moreover, heatmaps in Fig. 14(a) by our STS-HGCN-AL show that the

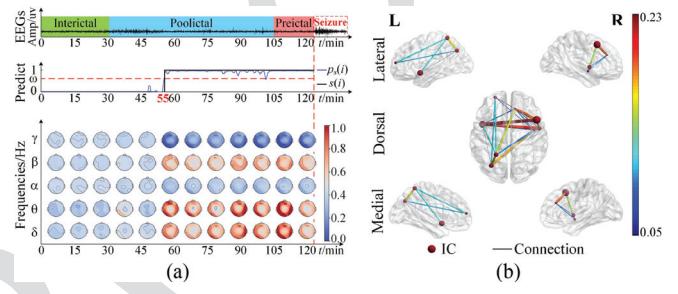


Fig. 14. (a) Visualization of the spectral-temporal dynamics for a 2 h duration prior to a specific seizure. (b) Spatial cortical dependencies in a specific patient (CHB 08).

frontal, temporal, and parietal regions are significantly critical to realize seizure prediction. To observe the spatially pathologic regions focused by the STS-HGCN-AL approach more precisely, we orient ICs via DIPFIT plugin of the EEGLAB toolbox with a boundary element model [41]. As a result, A_{ds} estimated by the GATENet is visualized in Fig. 14(b). Since each row in A_{ds} denotes the connections to the corresponding IC, node attributes in Fig. 14(b) are determined as the summation of the corresponding columns of A_{ds} ; whereas, edge properties are defined by the entities of A_{ds} directly. Only critical ICs with a node value of more than 1 are given in Fig. 14(b). We can see that critical IC pairs are mainly located in the frontal lobe, which is in line with the epileptogenic zone listed in the previous studies [42]. These results show that our STS-HGCN-AL can indeed capture spatio-temporal-spectral dynamics related to seizure initiation efficiently.

D. Performance Comparison of Different Seizure Prediction Methods Reported for CHB-MIT Database

Table VII summarizes the state-of-the-art seizure prediction methods on the CHB-MIT database, where NR is not reported values. Note that it is difficult to draw a direct comparison between our method and existing studies due to the different experimental settings, such as the selection of preictal and interictal data and validation scheme. From Table VII, noncontinuous preictal and interictal intervals exist in [2], [43], and [44]. Thus, the specificity of these studies,

that is, FDR/h, cannot be properly assessed due to the bias interictal selection, that is, noncontinuous selected preictal and interictal data [4]. Moreover, compared with our LOO cross-validation scheme, which does not break temporal continuity in the testing stage, k -fold CV in [5], [6], and [43] is much less challenging and practical since CV shuffles EEGs and ignores the intrapatient variation of seizures. As a result, our proposed STS-HGCN-AL scheme yields more promising S_n and FDR/h against most of the recent studies on the premise of testing on multiple patients with continuous EEG recordings using the LOOCV scheme. Note that it is not feasible to compare with over-optimistic results in [5] directly since their training and testing data in every CV loop may not be fully separated. Moreover, although rigorous statistical evaluation has been highly suggested [3], [4], only studies in [2], [43], and [45] have attempted to assess the statistical significance of their results against a chance predictor. On the contrary, our seizure predictor exceeds the chance level at all patients with $p < 0.05$.

E. Limitation and Future Directions

Although the proposed seizure predictor achieves a powerful seizure forecasting performance, two limitations still remain in the current work. First, although omitting preprocessing on EEG signals, artifact removal for example, can automatize the seizure prediction process as far as possible, the presence of redundant electrodes and artifacts in EEG data may alter the results. Thus, in future studies, we will focus on incorporating the adaptive EEG channel selection mechanism [46], [47], ant colony optimization-inspired optimal subset selection for example [48], and unsupervised artifact removal method [49] into the proposed STS-HGCN-AL to further boost the practical potential of our EEG seizure predictor. Second, although our method is superior to the state-of-the-art seizure prediction studies in a patient-specific scenario, namely, training and testing data are from the same patient, it cannot be used for nonpatient-specific seizure warning task directly, where the model is applied to a completely new patient after optimizing. This is mainly because our method does not own the abilities to handle the drifting distributions between training and test data. Thus, transfer learning strategies [50]–[53], common spatial pattern with domain adaptation [54] for instance, will be considered in our future work to fully promote our STS-HGCN-AL method.

VI. CONCLUSION

In this article, a novel EEG seizure predictor was proposed by means of STS-HGCN-AL. Specifically, our STS-HGCN-AL first captures temporal dynamics in an epileptic cortex under different rhythms via a subnet, called ST-SENet. Meanwhile, a variant self-gating unit, called GATENet, was adopted to infer ICs spatial couplings and prune spurious ones. Finally, multirhythm spatiotemporal properties were extracted and integrated alternatively by the proposed resGCN and rhythmAtt units, respectively. In addition, the proposed semisupervised active preictal interval learning combats the diversity of preictal transition among different patients

and, thus, aids our method to predict seizures robustly. Seizure prediction performance has been evaluated on multiple patients with abundant control recordings. Our method produces promising results in terms of AUC, S_n , and FDR/h. The efficacy of the temporal and spatial explorations by ST-SENet and GATENet was validated against handcrafted baselines, respectively. The effectiveness of the resGCN and rhythmAtt units was also evaluated against three counterparts. The potential application of our STS-HGCN-AL scheme, as an EEG seizure predictor, also relies on active preictal interval learning. Clinical investigation showed that our method can practically capture spatio-temporal-spectral dynamics related to seizure initiation. Experimental results illustrated that the proposed method can predict seizures efficiently and further aid epilepsy treatment and facilitate patients' daily routine life.

REFERENCES

- [1] L. Xie, Z. Deng, P. Xu, K.-S. Choi, and S. Wang, "Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2200–2214, Jun. 2019.
- [2] A. R. Ozcan and S. Erturk, "Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2284–2293, Nov. 2019.
- [3] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, Sep. 2007.
- [4] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction—Ready for a new era," *Nat. Rev. Neurol.*, vol. 14, no. 10, pp. 618–630, Oct. 2018.
- [5] K. M. Tsioris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, Aug. 2018.
- [6] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2109–2118, Sep. 2018.
- [7] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 804–813, Oct. 2019.
- [8] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial-temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.
- [9] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, Jul. 2020.
- [10] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affective Comput.*, early access, Aug. 27, 2019, doi: [10.1109/TAFFC.2019.2937768](https://doi.org/10.1109/TAFFC.2019.2937768).
- [11] Q. Lian, Y. Qi, G. Pan, and Y. Wang, "Learning graph in graph convolutional neural networks for robust seizure prediction," *J. Neural Eng.*, vol. 17, no. 3, Jun. 2020, Art. no. 035004, doi: [10.1088/1741-2552/ab909d](https://doi.org/10.1088/1741-2552/ab909d).
- [12] S. Jang, S. Moon, and J. Lee, "EEG-based video identification using graph signal modeling and graph convolutional neural network," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Calgary, AB, Canada, 2018, pp. 3066–3070.
- [13] M. Wang, H. El-Fiqi, J. Hu, and H. A. Abbass, "Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 3259–3272, 2019.
- [14] A. Coito, C. M. Michel, P. V. Mierlo, S. Vulliémoz, and G. Plomp, "Directed functional brain connectivity based on EEG source imaging: Methodology and application to temporal lobe epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 12, pp. 2619–2628, Dec. 2016.

- [15] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affective Comput.*, early access, May 11, 2020, doi: [10.1109/TAFFC.2020.2994159](https://doi.org/10.1109/TAFFC.2020.2994159).
- [16] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.
- [17] M. A. Kramer and S. S. Cash, "Epilepsy as a disorder of cortical network organization," *Neuroscientist*, vol. 18, no. 4, pp. 360–372, Aug. 2012.
- [18] F. Bartolomei *et al.*, "Defining epileptogenic networks: Contribution of SEEG and signal analysis," *Epilepsia*, vol. 58, no. 7, pp. 1131–1147, Jul. 2017.
- [19] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Aug. 2017.
- [20] G. Li, M. Müller, A. K. Thabet, and B. Ghanem, "DeepGCNs: Can GCNs go as deep as CNNs?" in *Proc. IEEE Int. Conf. Comput. Vis.*, Seoul, Korea, 2019, pp. 9266–9275.
- [21] C.-H. Yang, Y.-H. Shih, and H. Chiueh, "An 81.6 μ W FastICA processor for epileptic seizure detection," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 1, pp. 60–71, Feb. 2015.
- [22] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [23] S. M. Azimi, P. Fischer, M. Körner, and P. Reinartz, "Aerial LaneNet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2920–2938, May 2019.
- [24] Y. Li, W. G. Cui, H. Huang, Y. Z. Guo, K. Li, and T. Tan, "Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach," *Knowl. Based Syst.*, vol. 164, pp. 96–106, Jan. 2019.
- [25] F. G. Tang, Y. Liu, Y. Li, and Z.-W. Peng, "A unified multi-level spectral-temporal feature learning framework for patient-specific seizure onset detection in EEG signals," *Knowl. Based Syst.*, vol. 205, Oct. 2020, Art. no. 106152, doi: [10.1016/j.knosys.2020.106152](https://doi.org/10.1016/j.knosys.2020.106152).
- [26] K. P. Indiradevi, E. Elias, P. S. Sathidevi, S. D. Nayak, and K. Radhakrishnan, "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram," *Comput. Biol. Med.*, vol. 38, no. 7, pp. 805–816, Jul. 2008.
- [27] L. Wang *et al.*, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, p. 222, May 2017.
- [28] Y. Li, Y. Liu, W. G. Cui, Y. Z. Guo, H. Huang, and Z. Y. Hu, "Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 782–794, Apr. 2020.
- [29] P. J. Marshall, Y. Bar-Haim, and N. A. Fox, "Development of the EEG from 5 months to 4 years of age," *Clin. Neurophysiol.*, vol. 113, no. 8, pp. 1199–1208, Aug. 2002.
- [30] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Aug. 2017.
- [31] D. S. Bassett and E. Bullmore, "Small-world brain networks," *Neuroscientist*, vol. 12, no. 6, pp. 512–523, Dec. 2006.
- [32] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 6000–6010.
- [33] H. Wang, Y. Jin, and J. Doherty, "Committee-based active learning for surrogate-assisted particle swarm optimization of expensive problems," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2664–2677, Sep. 2017.
- [34] A. Chakrabarty, C. Danielson, S. D. Cairano, and A. Raghunathan, "Active learning for estimating reachable sets for systems with unknown dynamics," *IEEE Trans. Cybern.*, early access, Jul. 22, 2020, doi: [10.1109/TCYB.2020.3000966](https://doi.org/10.1109/TCYB.2020.3000966).
- [35] D. E. Snyder, J. Echauz, D. B. Grimes, and B. Litt, "The statistics of a practical seizure warning system," *J. Neural Eng.*, vol. 5, no. 4, pp. 392–401, Sep. 2008.
- [36] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Harvard-MIT Health Sci. Technol., Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
- [37] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup, "EMD-based temporal and spectral features for the classification of EEG signals using supervised learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 28–35, Jan. 2016.
- [38] A. Gupta, P. Singh, and M. Karlekar, "A novel signal modeling approach for classification of seizure and seizure-free EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 925–935, May 2018.
- [39] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, Palais des Congrès Neptune, Toulon, France, 2017, pp. 1–4.
- [40] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, pp. 818–833.
- [41] G. Chaitanya, S. Sinha, M. Narayanan, and P. Satishchandra, "Scalp high frequency oscillations (HFOs) in absence epilepsy: An independent component analysis (ICA) based approach," *Epilepsy Res.*, vol. 115, pp. 133–140, Sep. 2015.
- [42] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 465–474, Aug. 2020.
- [43] D. Cho, B. Min, J. Kim, and B. Lee, "EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1309–1318, Aug. 2017.
- [44] N. D. Truong *et al.*, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw.*, vol. 105, pp. 104–111, Sep. 2018.
- [45] A. S. Zandi, R. Tafreshi, M. Javidan, and G. A. Dumont, "Predicting epileptic seizures in scalp EEG based on a variational bayesian gaussian mixture model of zero-crossing intervals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1401–1413, May 2013.
- [46] F. Qi *et al.*, "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, Feb. 2021.
- [47] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.
- [48] X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Appl. Soft Comput.*, vol. 24, pp. 585–596, Nov. 2014.
- [49] A. Mur, R. Dormido, and N. Duro, "An unsupervised method for artefact removal in EEG signals," *Sensors*, vol. 19, no. 10, p. 2302, May 2019.
- [50] T. Zhang, G. Su, C. Qing, X. Xu, B. Cai, and X. Xing, "Hierarchical lifelong learning by sharing representations and integrating hypothesis," *IEEE Trans. Syst. Man. Cybern., Syst.*, vol. 51, no. 2, pp. 1004–1014, Feb. 2021.
- [51] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and dempster-shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 24, 2020, doi: [10.1109/TNNLS.2020.3015505](https://doi.org/10.1109/TNNLS.2020.3015505).
- [52] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.
- [53] X. Li *et al.*, "EEG-based mild depression recognition using convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 57, no. 6, pp. 1341–1352, Jun. 2019.
- [54] B. Wang *et al.*, "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Trans. Cybern.*, early access, Apr. 22, 2020, doi: [10.1109/TCYB.2020.2982901](https://doi.org/10.1109/TCYB.2020.2982901).



Yang Li received the Ph.D. degree in automatic control and systems engineering from the University of Sheffield, Sheffield, U.K., in 2011.

He did Postdoctoral Research with the Department of Computer and Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, for one year. In 2013, he joined the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing, China, as an Associate Professor, and was promoted to a Professor in 2018.

His current research interests include system identification and modeling for complex nonlinear processes: NARMAX methodology and applications, nonlinear and nonstationary signal processing, intelligent computation and data mining, medical image processing, and deep learning.

1311
1312
1313
1314
1315
1316
1317
1318
1319



Yu Liu received the B.Eng. degree in automation from the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing, China, in 2018, where she is currently pursuing the M.Eng. degree in pattern recognition and intelligent systems.

Her research interests include biomedical imaging and signal analysis, epileptic seizure detection and prediction, machine learning, and deep learning.

1320
1321
1322
1323
1324
1325
1326
1327
1328
1329



Yu-Zhu Guo received the B.Sc. and M.Sc. degrees from the Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in automatic control and systems engineering from The University of Sheffield, Sheffield, U.K., in 2009.

He is currently an Associate Professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing. His research interests include system identification and information processing for nonlinear systems, NARMAX methods, nonlinear spectral analysis, spatiotemporal systems, multiscale modeling of neuromusculoskeletal systems, brain-inspired intelligence and computing, and bioinformatics.

1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343



Xiao-Feng Liao (Fellow, IEEE) received the B.S. and M.S. degrees in mathematics from Sichuan University, Chengdu, China, in 1986 and 1992, respectively, and the Ph.D. degree in circuits and systems from the University of Electronic Science and Technology of China, Chengdu, China, in 1997.

He was a Postdoctoral Researcher with Chongqing University, Chongqing, China, from 1999 to 2001. He was a Research Associate with the Chinese University of Hong Kong, Hong Kong, from 1997 to 1998, where he was a Research Associate from 1999 to 2000, a Senior Research Associate from 2001 to 2002, and a Research Fellow from 2006 to 2007. From 1999 to 2012, he was a Professor with Chongqing University. He is currently a Professor with Chongqing University, where he is the Dean of College of Computer Science and Engineering. His current research interests include neural networks, nonlinear dynamical systems, bifurcation and chaos, and cryptography.



Bin Hu (Senior Member, IEEE) received the M.S. degree in computer science from the Beijing University of Technology, Beijing, China, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science, Beijing.

From 2007 to 2009, he was a Reader with the Head of Context Aware Computing Research Group, School of CTN, Birmingham City University, Birmingham, U.K. Since 2009, he has been the Dean of the School of Information Science and

Engineering, Lanzhou University, Lanzhou, China. He is a Guest Professor with ETH Zurich, Zürich, Switzerland. His current research interests include pervasive computing, cognitive computing, and mental healthcare.

Prof. Hu served as an Editor for *IET Communications, Cluster Computing, Wireless Communications and Mobile Computing*, the *Journal of Internet Technology, Security and Communication Networks* (Wiley), and *Brain Informatics*, and an Associate Editor of some peer-reviewed journals in computer science, such as the *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING* and the *IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS*. He is also an IET Fellow, an IET Fellow Assessment Panel Member (China Committee), the Co-Chair of IEEE SMC TC on Cognitive Computing, the Member-at-Large of ACM China, the Director of the Web Intelligence Consortium (China Committee), and a Board Member of the International Society for Social Neuroscience (China Committee).



Tao Yu received the M.D. and Ph.D. degrees from the Capital Medical University, Beijing, China, in 2006.

He is currently a Professor with the Department of Functional Neurosurgery, Xuanwu Hospital, Capital Medical University, and the Vice Director with the Beijing Institute of Functional Neurosurgery, Beijing. His research interests include presurgical evaluation of refractory epilepsy, neuromodulation of refractory epilepsy, and functional mapping of the cortex.